

ECOLE NATIONALE VETERINAIRE DE LYON

Année Scolaire 1939-1940 — No 2

Contribution à l'étude DE LA BIOMETRIE

THÈSE

PRÉSENTÉE

A LA FACULTÉ DE MÉDECINE ET DE PHARMACIE DE LYON

et soutenue publiquement le 22 Décembre 1939

POUR OBTENIR LE GRADE DE DOCTEUR VÉTÉRINAIRE

PAR

Louis MEYER

Né le 18 Juillet 1917, à BUHL (Haut-Rhin)



LYON

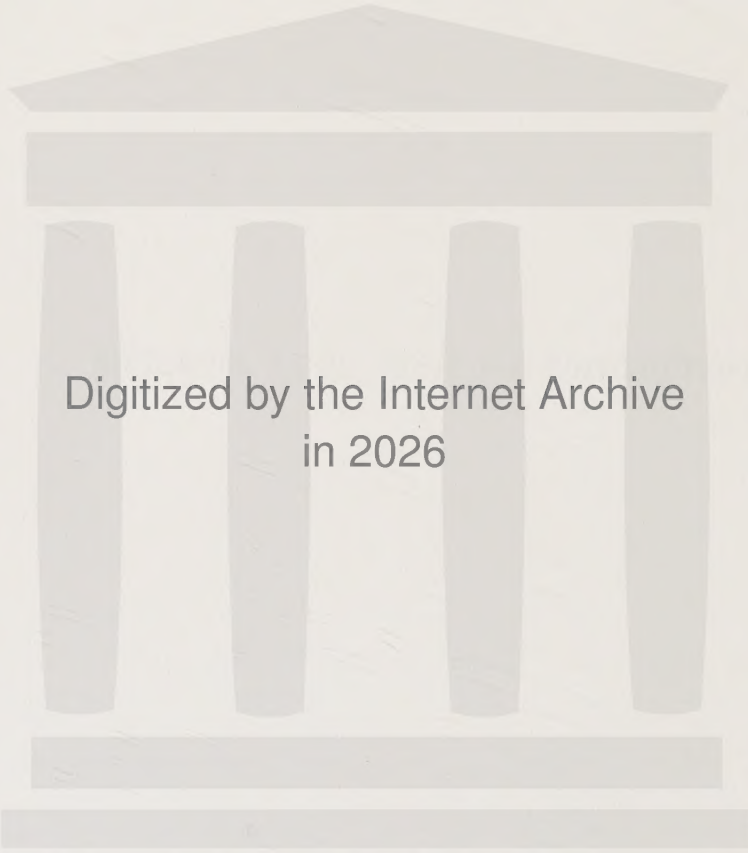
BOÛC Frères M. & L. RIOU

IMPRIMEURS-ÉDITEURS

42, Quai Gailleton, 42

—
1939

CONTRIBUTION A L'ÉTUDE DE LA BIOMETRIE



Digitized by the Internet Archive
in 2026

ECOLE NATIONALE VETERINAIRE DE LYON

Année Scolaire 1939-1940 --- No 2

Contribution à l'étude DE LA BIOMÉTRIE

THÈSE

PRÉSENTÉE

A LA FACULTÉ DE MÉDECINE ET DE PHARMACIE DE LYON

et soutenue publiquement le 22 Décembre 1939

POUR OBTENIR LE GRADE DE DOCTEUR VÉTÉRINAIRE

PAR

Louis MEYER

Né le 18 Juillet 1917, à BUHL (Haut-Rhin)



LYON

BOSC Frères M. & L. RIOU

IMPRIMEURS-ÉDITEURS

42, Quai Gailleton, 42

—
1939

PERSONNEL ENSEIGNANT DE L'ECOLE VÉTÉRINAIRE DE LYON

Directeur M. JUNG.
 Directeur honoraire ... M. F.-X LESBRE.
 Professeurs honoraires. M. BALL.
 M. CADEAC.

PROFESSEURS

Physique, Chimie et Toxicologie, Pharmacie.....	MM. TAPERNOUX.
Anatomie descriptive des animaux domestiques, Tératologie	TAGAND.
Physiologie, Thérapeutique générale, Matière médi- cale	JUNG.
Histologie et Embryologie, Anatomie pathologique, Inspection des viandes.....	BALL.
Parasitologie et maladies parasitaires, Zoologie ap- pliquée, Botanique, Clinique des maladies pa- rasitaires	MAROETL.
Pathologie médicale des Equidés et des Carnassiers, Séméiologie et Propédeutique, Jurisprudence vétérinaire, Médecine légale, Clinique médicale.	BRION.
Pathologie chirurgicale, Séméiologie, Médecine opé- ratoire et ferrure, Clinique chirurgicale.....	DOUVILLE.
Pathologie bovine, ovine, porcine, caprine, Maladies des oiseaux de basse-cour, Séméiologie, Méde- cine opératoire, Obstétrique, Cliniques spéciales	CUNY.
Maladies microbiennes, Police sanitaire, Microbio- logie et Pathologie générale, Clinique spéciale..	COLLET.
Zootechnie et appréciation des animaux domestiques, Aliments du bétail, Agronomie et Hygiène des animaux, Economie rurale.....	JEAN-BLAIN.

CHEFS DE TRAVAUX

MM. COLLET, Agrégé
 GRATECOS, MADELENAT et FERRANDO

EXAMINATEURS DE LA THÈSE

Président : M. le Professeur ARLOING.
1^{er} Assesseur : M. le Professeur DOUVILLE.
2^{me} Assesseur : M. le Professeur BALL.



La Faculté de Médecine et l'Ecole Vétérinaire déclarent que les
 opinions émises dans les dissertations qui leur sont présentées doivent
 être considérées comme propres à leurs auteurs et qu'elles n'entendent
 leur donner ni approbation ni improbation.

A MON PERE ET A MA MERE

Que ce travail soit, avec l'assurance de mon amour filial, le témoignage de ma reconnaissance pour leur dévouement et les nombreux sacrifices qu'ils se sont imposés en vue de mon instruction et celle de mon frère.

A MON FRERE

Qu'il soit sûr de trouver toujours en moi plus qu'un ami fidèle et dévoué.

A TOUTE MA FAMILLE

A TOUS MES AMIS

A MONSIEUR LE PROFESSEUR ARLOING

De la Faculté de Médecine de Lyon
Professeur de Médecine Expérimentale et Comparée
et Bactériologie
Chevalier de la Légion d'Honneur

Qui nous a fait le grand hon-
neur d'accepter la présidence du
Jury de notre thèse. Nous l'en
remercions bien vivement.

A MONSIEUR LE PROFESSEUR JEAN-BLAIN

Qui fut pour nous un maître et
un conseiller. Qu'il trouve ici l'ex-
pression de notre sincère grati-
tude.

A MONSIEUR LE PROFESSEUR DOUVILLE

Tous nos remerciements pour
l'enseignement qu'il nous a donné
et la bienveillance qu'il nous a
témoignée.

A TOUS NOS MAÎTRES

DE L'ECOLE NATIONALE VÉTÉRINAIRE DE LYON

DE L'ECOLE MILITAIRE ET D'APPLICATION

DE LA CAVALERIE ET DU TRAIN

CONTRIBUTION A L'ÉTUDE DE LA BIOMETRIE

AVANT-PROPOS

Il est actuellement, en Zootechnie Générale, une question de plus en plus étudiée, mais encore assez discutée : la valeur des corrélations; elle forme une partie importante de la Biométrie, science de connaissance relativement récente.

Aussi, lorsque, hésitant sur l'entreprise de cette thèse, nous avons demandé conseil à notre professeur de Zootechnie, M. Jean-Blain, il nous confirma immédiatement dans cette voie, ce dont nous lui sommes actuellement très reconnaissant, car c'est un sujet qui nous a vivement intéressé et dont nous suivrons les progrès même lorsque nous ne pourrons plus prendre notre faible part au débat, lorsque nous serons le « praticien-bohème » de quelque clientèle de campagne.

Qu'il nous soit donc permis d'exprimer notre gratitude à M. le Professeur Jean-Blain pour ses conseils et son appui dévoués, ainsi qu'à M. le Commandant Douville, qui a bien voulu prendre à sa charge le remplacement de M. Jean-Blain, obligé de s'absenter.

Nos remerciements vont aussi à M. le Professeur Arloing, de la Faculté de Médecine de Lyon qui nous

a fait le grand honneur d'accepter la présidence de notre Jury.

Enfin nous prions tous nos Maîtres de l'Ecole de Lyon, à qui nous sommes redevable de notre succès dans les études, d'agréer ici l'hommage de notre reconnaissance et de notre respect.

INTRODUCTION

La Biométrie est l'estimation méthodique, mathématique, des données pouvant être mesurées ou pesées, fournies par le corps animal.

Les méthodes modernes nous sont actuellement d'une aide précieuse dans l'étude de la complexité de l'être vivant; elles nous permettent une représentation exacte de cette complexité, mais non, comme nous le verrons plus loin, son explication. Ici, comme partout ailleurs, les recherches purement empiriques ont dû céder le pas à des méthodes modernes qui sont basées sur des moyens scientifiques. Nous avons à notre disposition des moyens mathématiques tels que les méthodes binomiale, arithmétique et graphique qui permettent une estimation comparative très exacte.

Tout le but de la Biométrie repose sur l'étude des variations et, comme nous l'avons déjà fait remarquer, le chapitre des corrélations en forme la partie la plus ardue et la plus discutée; il constitue en quelque sorte, s'il nous est permis d'employer ce cliché, « la bouteille à encre » de la Biométrie. Mais avant d'arriver à cette partie si importante, il nous faudra d'abord étudier les variations en elles-mêmes et en particulier leur étendue, leur moyenne, la déviation

standard et l'erreur. Pour chacun de ces termes nous donnerons la définition, des exemples fournis par nos recherches personnelles, et enfin leur application pratique.

Après avoir vu les variations en elles-mêmes nous passerons à la seconde partie de ce travail, qui sera constituée par l'étude des relations entre elles, c'est-à-dire l'étude des variations. Nous envisagerons successivement leur base théorique, les applications que nous en avons faites personnellement et enfin l'interprétation biologique et la critique.

Nos recherches personnelles ont été faites au marché et dans les étables des Abattoirs de la Mouche. Parmi les nombreuses et diverses races présentées, nous avons choisi la race Limousine et tout particulièrement dans cette race le sujet préparé pour la boucherie et prêt à être abattu; il est représenté par des taurillons et des génisses sans dents de remplacement mais dont les pinces commencent déjà à se déchausser, soit donc aux environs de 18 mois. Ces animaux, de pure race Limousine, ont environ 1 m. 15 de hauteur au garrot et ont des cornillons de 4 à 6 cm. de long; leur état d'engraissement est très bon, c'est le sujet idéal pour la boucherie.

Nous avons préféré cette race pour quelques raisons d'ordre théorique et pratique. Au point de vue théorique il nous a semblé utile, pour pouvoir étudier la valeur des méthodes biométriques, de ne nous adresser qu'à une race pure : la race limousine est une de celles qu'on a le plus de facilités à observer à l'état de pureté absolue. En prenant un croisement,

ou simplement une catégorie d'animaux renfermant un tant soit peu de sang étranger, nous serions obligés de prendre en considération de nombreux facteurs différents, ce qui ne nous permettrait pas d'arriver à notre but. Au point de vue pratique, c'est leur nombre, leur uniformité de format et leur docilité — permettant plus de précision dans les mesures — qui nous ont tentés.

Nous avons relevé les variations de six dimensions :

- H : la hauteur au garrot;
- L : la longueur de la tête, mesurée du bout du mufle au milieu du chignon;
- l : la largeur de la tête prise au-dessus des orbites;
- h : l'écartement des hanches : distance entre les angles externes de l'ilium;
- i : l'écartement des ischiums : pris au niveau des tubérosités ischiales;
- l.c. : la longueur de croupe : distance séparant l'angle externe de l'ilium de la tubérosité ischiale.

Le nombre des bovins examinés a été de 130.

Nous ne donnerons pas ici les résultats in extenso, ce qui serait trop fastidieux, mais nous les énoncerons au fur et à mesure de leurs applications.

LA VARIABILITE

En examinant un groupement d'individus de même espèce et au même endroit, groupement que l'on appelle une population, nous constatons que, même dans le cadre d'une parenté très étroite, il n'y a jamais similitude absolue entre tous les constituants. Nous trouvons toujours des « variations » ou une « variabilité », si petite soit-elle.

Ainsi, dans nos mesures de longueur de tête des 130 limousins, nous avons trouvé les chiffres L suivants; sous chaque valeur de L nous portons le nombre n d'individus qui les présentaient :

L	37...	39	40	41	42	43	44	45	46	47	48...	51	52
n	1	2	6	12	18	24	30	17	12	4	2	1	1

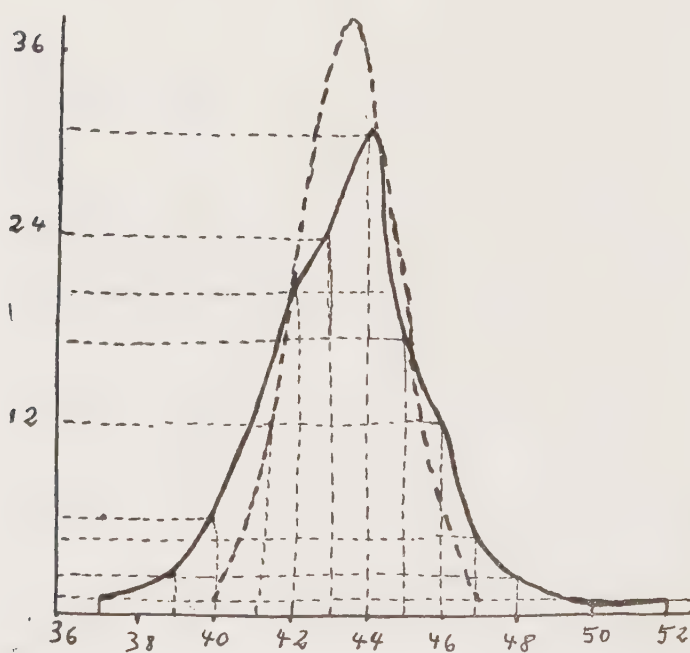
Cette énumération est nommée « suite de variations » ou plus simplement « variation » et le nombre n est la « fréquence ».

Ces premières définitions posées, voyons maintenant ce qu'il faut entendre par « étendue de la variation ».

Elle est encore appelée, selon les auteurs, « latitude de variabilité » (Kronacker) ou « écart absolu » (auteurs belges).

Elle est déterminée mathématiquement par la différence qui existe entre la valeur de la variante la

plus élevée et celle de la variante la plus faible. Elle indique donc le maximum d'écart existant entre deux variantes, ce qui lui a valu la dénomination d'écart absolu; c'est dans les limites qu'elle indique que se trouvent toutes les variantes, d'où l'autre synonymie : latitude de variabilité. Ainsi, dans l'exemple précédent, l'étendue de la variation est de 52 cm — 37 cm. = 15 cm.



La suite de variations peut encore être représentée sous une forme plus schématique que l'énumération : c'est la méthode graphique : on porte en ordonnées le nombre de sujets ayant une même variante et en abscisses, la série des variantes; nous obtenons une

figure qui porte le nom de « courbe de fréquence ». La courbe suivante (en trait plein) est celle de la longueur de tête.

L'étendue de la variation y est représentée par la base de la courbe qui est la partie de l'abscisse comprise entre 37 et 52 cm.

La mesure de la variabilité, ainsi que sa représentation graphique sont, pour les zootechniciens, d'une certaine utilité. Elles leur permettent de se faire une première idée sur la fixité et mieux encore sur la pureté d'une population, d'une race. Plus une race sera pure, plus son type sera uniforme et par conséquent il y aura le moins d'écart possible entre les extrêmes : l'étendue de la variabilité doit être réduite à un minimum.

Un exemple frappant nous est donné par les mesures prises par certains auteurs sur différentes races étrangères. Ainsi, pour la hauteur au garrot, W. Zorn donne pour la race des montagnes de la région de Glatz 19 cm., L. Zehntner pour le bétail rouge de Silésie 22 cm, pour la Simmenthal 26 cm.

L'interprétation de ces résultats est facile : le bétail de la région de Glatz, qui présente le minimum de variabilité, est un bétail très localisé, dans une région montagneuse, à communications difficiles, donc peu sujet à des modifications apportées des environs, mais par contre soumis et adapté à un climat et des circonstances de vie très spéciales, ce qui a contribué à son unification. Au contraire, la Simmenthal présente un type de bovins très répandu dont le commerce est actif et dont le mélange des différentes souches est

fréquent; on comprend aisément une moins grande homogénéité. Mais que penser alors de notre race française du Limousin, qui passe pour être une race sélectionnée et fixée et pour laquelle la hauteur au garrot varie de 103 cm. à 132 cm., soit de 29 cm. ? Faut-il en conclure qu'elle n'est pas pure ? Ce serait nous avancer un peu loin, car, nous le verrons dans la suite, cette appréciation d'une race par la seule valeur de l'écart absolu est sujette à quelques critiques. Il ne faut pas oublier non plus qu'il s'agit d'individus non adultes, mais à un âge où les variations sont bien plus étendues par suite des variations de croissance, très influencées par des facteurs individuels ou collectifs.

De même l'examen du graphique des fréquences permet de constater l'amplitude de la variabilité. Cette conclusion se déduit de l'allure générale de la courbe. On peut dire que plus la courbe est haute par rapport à sa largeur, plus le type est uniforme; en effet, dans ce cas, à une base très petite — ce qui signifie un écart absolu petit — correspond une hauteur très grande — donc un très grand nombre d'individus. Au contraire, plus elle est basse par rapport à sa largeur, plus la variation du groupe envisagé est grande.

Le graphique représenté précédemment nous donne le type d'une courbe de variabilité moyenne. L'idéal, la perfection serait évidemment la courbe en pointillé qui est obtenue par la représentation graphique du développement du binôme $(a + a)^7$ qui, pour un

total de 128 unités, les répartit comme suit :
1—7—21—35—35—21—7—1—.

Nous pourrions toutefois constater que les deux courbes ne présentent pas de différences trop grandes. Assez uniforme, notre courbe n'a qu'un seul sommet, elle est dite « unimodiale », mais il existe aussi des courbes bi- et même pluri-modiales; elles se rencontrent lorsqu'on opère en dehors des lignées pures, dans les populations mélangées. Ceci n'est toutefois pas une règle absolue et même dans une lignée pure il arrive d'obtenir des courbes bi-modiales par suite de la prédominance variable d'un caractère ethnique important.

La comparaison de suites de variations ou de graphiques peut encore servir à la recherche de l'influence d'un facteur donné sur une constitution fixe. Pour cette étude on s'adressera par exemple à des populations de même race, voisines pour que les influences climatiques et les ressources agraires soient les mêmes; on modifiera un facteur quelconque — soit par amendement du sol, soit par changement du régime — et on établira les variations. Sur ce point nos recherches personnelles ont été nulles, car les bovins que nous avions à notre disposition provenaient tous de la même région et leur régime nous était inconnu.

Nous sommes donc obligés de citer les conclusions de Kronacker ayant expérimenté sur des Simmenthal de deux régions ne différant, dit-il, que par leur humidité. « Par la comparaison des pourcentages des « fréquences des différentes classes, on voit par des

« chiffres l'influence d'un caractère dominant et différentiel d'une région, dans notre cas, l'humidité. »

Mais cet écart absolu n'est pas parfait pour juger de la variabilité d'un type. On peut lui faire le grand reproche suivant : il varie avec le nombre des sujets examinés. Si nous prenons 64 fréquences qui se disposent, dans le cas théorique, suivant le développement du binôme $(a + a)^6$ on aura comme fréquences 1—6—15—20—15—6—1 correspondant à des classes $X + 2$, $X + 3$, $X + 4$, $X + 5$, $X + 6$, $X + 7$: l'écart absolu sera de $(X + 7) - (X + 1) = 6$. Si, au lieu d'étudier 64 individus nous en étudions 256, les fréquences constituent le binôme $(a + a)^8$ et l'écart absolu sera 8, ce qui ne veut pas dire que la population des 256 sujets est moins pure que celle des 64, car en même temps que des extrêmes plus éloignées nous aurons un nombre infiniment plus grand de variations centrales qui, par leur peu de variabilité, viendront neutraliser les effets des deux nouvelles extrêmes.

Pour arriver à déterminer l'écart absolu réel fixe, celui qui pour une race n'est plus capable d'augmenter, il faut un très grand nombre d'observations, parfois de l'ordre de 10.000 et même plus. On ne sait donc jamais, ou au moins que rarement, si les mesures prises sont assez nombreuses pour exclure la possibilité de l'existence de variantes extrêmes plus basses ou plus élevées.

C'est pourquoi nous substituerons bientôt à l'écart absolu deux autres valeurs, moins sujettes à variation, l'écart moyen et la déviation standard.

Enfin, avant de quitter ce chapitre de généralités

sur la variabilité, disons que les règles de la variabilité, dites lois de Quetelet, ne sont applicables qu'avec réserves à des populations sélectionnées et surtout en voie de sélection.

En effet, les petites différences existant entre les individus d'une population proviennent, pour Blaringhem, de la réaction à une multitude de petits facteurs indépendants les uns des autres et tendant à former des combinaisons variées qui ont toutes un nombre égal de chances de se produire. En partant de cette hypothèse on comprend qu'il suffit de permettre à l'une de ces petites influences de se développer plus que ses congénères pour constater une altération aux lois des variations.

Auparavant existait l'uniformité des degrés et l'équilibre des formes comme on les retrouve encore chez les animaux sauvages; après cette action de l'homme, la sélection, il reste un déséquilibre voulu, entretenu, un isolement des groupes extrêmes suivant le but et l'usage que veut en faire le zootechnicien; ce dernier devra s'en souvenir dans ses expériences.

Les exemples ne manquent pas; il suffit de suivre l'historique de n'importe quelle race sélectionnée; prenons le premier cas venu, l'amélioration de la race ovine Southdown.

Au début, dans le comté du Sussex, on trouvait une race rustique et sobre : le Southdown primitif; la sélection vitale jouait seule, le troupeau était bien équilibré, les variantes extrêmes étaient rares. Puis intervinrent John Ellemann et Johnas Bern : ils choi-

sirent les sujets les mieux conformés et les plus lourds qui, seuls, eurent le droit et le devoir d'être des reproducteurs : il y eut donc à ce moment un élargissement de la courbe des fréquences, l'écart absolu augmenta, l'allure de la courbe fut momentanément modifiée : la sélection avait faussé l'équilibre établi par la nature.

Mais, par suite de la continuité de la sélection et de la fixation des caractères, ce déséquilibre tend à diminuer et, la race une fois fixée, nous aurons de nouveau une courbe binomiale et monomodiale mais avec des variantes extrêmes plus capricieuses par suite des cas d'atavisme se produisant brusquement; plus ces derniers seront rares, plus la race sera fixée.

Cette évolution, artificielle pour les animaux domestiques, mais naturelle et plus faible pour les espèces sauvages, n'est pas illimitée, comme on l'a vu, mais au contraire nettement limitée génotypiquement, chaque caractère spécifique ou ethnique ne pouvant évoluer que dans les limites des possibilités de dissociation et de recomposition des composants de son génotype. Cette évolution est aussi limitée phénotypiquement, chacun de ces caractères nouveaux ne pouvant présenter que des modifications bien déterminées, sous l'influence de milieux nettement différents. Les races semblent être les termes ultimes de la grande évolution animale dans laquelle les espèces correspondantes représenteraient des échelons moins évolués parce qu'encore plus complexes.

Nous n'irons donc pas jusqu'à prétendre, avec Blaringhem, que les lois de la variabilité ne sont vala-

bles que pour des populations non sélectionnées; nous dirons seulement qu'elles se présentent très exactes dans ce dernier cas, mais aussi, bien que plus sujettes à des exceptions sans explications apparentes, dans les races sélectionnées; par contre elles sont absolument inapplicables aux espèces en voie de sélection.

MOYENNE DES VARIATIONS

Dans une suite de variations les fréquences débutent par des chiffres faibles, augmentent plus ou moins régulièrement pour atteindre un maximum, puis décroissent de façon symétrique à la croissance. Les variantes supérieures et inférieures semblent réparties également autour d'un nombre appelé moyenne.

Reprenons l'exemple : variation de la largeur de tête :

l :	15	16	17	18	19	20	21	22
n :	2	11	24	34	35	20	3	1

n croît de 2 jusqu'à la moyenne de 34-35, puis décroît, mais la courbe décroissante est un peu plus inclinée que la courbe croissante. Néanmoins on a ici cette sorte d'équilibre partiel que nous avons déjà mentionné, entre les variantes supérieures et inférieures.

Pour une lignée pure on doit toujours obtenir cette répartition, à condition bien entendu de s'adresser à un nombre suffisant de sujets. Quetelet a démontré que cette disposition se fait très approximativement suivant la formule du binôme $(a + a)^x$. La suite des coefficients obtenus mathématiquement doit s'approcher de celle des différentes classes obtenues expéri-

mentalement. Dans ces développements, X étant le coefficient du binôme, le groupement des fréquences, c'est-à-dire le nombre des classes sera $X+1$ et la somme des fréquences, donc le nombre total des variations sera la somme des coefficients de a soit $n = (1+1)^X = 2^X$, comme il est possible de s'en convaincre à première vue par le tableau suivant. Quant à la répartition des fréquences, elle se fait suivant un dispositif spécial dit triangle de Pascal, que voici :

$(a+a)^0$	1	1
$(a+a)^1$	1 + 1	2
$(a+a)^2$	1 + 2 + 1	4
$(a+a)^3$	1 + 3 + 3 + 1	8
$(a+a)^4$	1 + 4 + 6 + 4 + 1	16
$(a+a)^5$	1 + 5 + 10 + 10 + 5 + 1	32
$(a+a)^6$	1 + 6 + 15 + 20 + 15 + 6 + 1	64
$(a+a)^7$	1 + 7 + 21 + 35 + 35 + 21 + 7 + 1	128

Pour nos mesures qui portent sur 130 unités il faudrait obtenir une forme approchant $(a+a)^7 = 1-7-21-35-35-21-7-1$, dont la somme est 128. Ce sont là les lois de Quetelet ou loi des variabilités fluctuantes.

Ce que nous avons utilisé jusqu'à présent comme moyenne est le chiffre le plus élevé des fréquences, et nous n'avons que supposé qu'il correspondait avec elle. Il existe une moyenne vraie, moyenne arithmétique. Pour la calculer deux procédés sont à notre disposition.

A. Le premier est basé sur le calcul ordinaire de la moyenne arithmétique de deux chiffres a et b ;

$M = \frac{a+b}{2}$. Soit $d_1, d_2, d_3, \dots$ les variantes et

$f_1, f_2, f_3, \dots$ les fréquences correspondantes, n leur somme; la moyenne sera

$$M = \frac{d_1 f_1 + d_2 f_2 + d_3 f_3 \dots}{n} = \Sigma \frac{d f}{n}$$

Comme application pratique calculons la moyenne de l'écartement des ischiurs de nos limousins. Nous avons obtenu la variation suivante :

i	23	24	25	26	27	28	29	30	31	32
f	5	15	11	28	31	22	10	3	3	2
$\Sigma f. i$	$= 115 + 360 + 275 + 728 + 837 + 616 + 290 + 90 + 93 + 64$									

$$M = \frac{\Sigma f i}{n} = \frac{3468}{130} = 26,68.$$

La moyenne arithmétique de l'écartement des ischiurs chez cette population est donc comprise entre 26 et 27 et est bien proche de la moyenne supposée 27.

B. Ce procédé, très exact, a l'inconvénient d'être un peu long pour des grandes suites de variations et l'on a plutôt recours au suivant qui, tout aussi exact, permet un calcul plus rapide. Il est basé sur le raisonnement suivant :

On choisit entre les variantes un chiffre A nommé moyenne supposée; en pratique ce sera la variante correspondant au nombre de fréquences le plus élevé: Si ce chiffre A correspond avec la moyenne vraie M, il faut que les déviations des autres variantes soient égales à partir d'elle, vers la gauche et vers

la droite, tout comme dans le cas théorique binomial. Au contraire, si les déviations positives (vers la droite) l'emportent sur les négatives (vers la gauche), la moyenne supposée est à gauche de la moyenne vraie et il faudra faire une rectification vers la droite ou inversement. Le chiffre de correction que nous représenterons par la lettre b est obtenu en divisant l'excès de l'une des déviations par le nombre total des individus.

On l'affectera du signe $+$ si les positifs l'emportent et du signe $-$ si les négatifs l'emportent. On aura donc $M = A \pm b$.

En pratique on numérotera les classes de variantes à partir de la moyenne supposée A en $+1+2+3...$ vers la droite et en $-1-2-3$ vers la gauche.

Calculons la moyenne de nos hauteurs au garrot :

H	103.	104.	105.	106.	107.	108.	109.	110.	111.	112.	113.	114.	115.	116.
H'	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
n	1	1		1		2	1	4	2	7	9	13	14	18

117.	118.	119.	120.	121.	122.	123.	124.	125.	126.	127.	132.
+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+16
14	10	9	4	6	4	4	2		2	1	1

La moyenne supposée est 116 cm. avec 18, chiffre le plus élevé, pour fréquence. Nous établirons le tableau suivant où nous porterons d'abord le numéro de déviation a et pour chaque valeur de a les fréquences positives et négatives; nous ferons la différence qui sera multipliée par la grandeur de la déviation a . Enfin il nous suffira de diviser la somme

de ces produits par le nombre n de fréquences
 $n = 130$ pour obtenir l'élément correctif b .

Déviati	a	1	2	3	4	5	6	7	8	9	10	11	12	13	16
fréquence +		14	10	9	4	6	4	4	2	2	1				1
fréquence —		14	13	9	7	2	4	1	2	1			1	1	
Différence		0	—3	0	—3	+4	0	+3	0	+1	+1	—1	—1	+1	
Diff. $\times a$		0	—6	0	—12	+20	0	+21	0	+10	+11	—12	—13	+16	
Σ diff. $\times a$		= (20 + 21 + 10 + 11 + 16) — (6 + 12 + 12 + 13) = 78 — 43 = 35.													

$$\text{Par conséquent } b = + \frac{35}{130} = + 0,27$$

La correction à apporter est donc de + 0,27.

On aura $M = A + 0,27 = 116 + 0,27 = 116,27$ cm.

La moyenne exacte est d'onc bien proche de la moyenne supposée, ce qui prouve que malgré l'étendue de la variation et son apparente irrégularité il y a tout de même un certain équilibre autour du maximum.

Certains auteurs utilisent encore une autre donnée, nommée Médiane. C'est la valeur du terme central d'une série de variantes ; elle a pour formule

$$\text{générale : } Me = \frac{n+1}{2} \text{ mais à notre avis cette donnée}$$

n'a pas une grande valeur, car elle est vraiment trop modifiable par une extrême parfois complètement isolée telle que la dimension 132, correspondant à la classe 16 éloignée de 5 de sa voisine. De plus, cette Médiane ne correspond à rien d'effectif ni de pratique.

Dans le chapitre précédent nous avons critiqué l'écart absolu par suite de sa variabilité avec le nom-

bre des observations. L'écart moyen subit aussi un peu ce reproche mais dans une moindre mesure. Cet écart moyen est un élément de calcul qui est obtenu

par la formule $\Sigma \frac{a f}{n}$, a représentant, comme dans le cas précédent, l'écartement d'une classe de la moyenne, f la fréquence et n le nombre total des fréquences.

Nous calculerons l'écart moyen de la variation suivante, variation de la longueur de croupe de nos limousins :

a	—6	—5	—4	—3	—2	—1	0	+1	+2	+3	+4	+5	+6	+7
l c	35	36	37	38	39	40	41	42	43	44	45	46	47	48
f	1	4	10	16	13	17	17	15	16	11	4	3	2	1

Il faudra d'abord que nous cherchions la moyenne :
 $M = A \pm b = 40,8$.

Pour la commodité du calcul nous arrondirons à 41, ce qui représente $f = 17$. A partir de ce point nous numérotions les écarts et nous recherchons la somme des produits fréquence par écart sans nous occuper des signes :

$$\Sigma a. f = 15.1 + 16.2 + 11.3 + 4.4 + 3.5 + 2.6 + 1.7 + 17.1 \\ + 13.2 + 16.3 + 10.4 + 4.5 + 1.6 = 287.$$

$$\text{L'écart moyen } e = \frac{\Sigma a f}{n} = \frac{287}{130} = 2,2.$$

De même qu'en ajoutant ou en retranchant de la médiane la moitié de l'écart absolu, on obtient les extrêmes, de même en ajoutant et en retranchant de la moyenne l'écart moyen on obtient deux variantes qui sont nommées, suivant le cas, variante pro-

bable inférieure ou supérieure et c'est entre ces deux variantes qu'on aura le plus de chance de placer une mensuration isolée. Mais de même que l'écart moyen, ces extrêmes probables ne présentent qu'un intérêt minime, car elles restent trop théoriques et n'ont pas grande signification.

L'écart moyen lui-même se modifie par les mêmes facteurs que l'écart absolu, mais, il est vrai, dans une mesure moindre. De plus, il arrive qu'il soit le même pour deux variations pourtant différentes ; c'est pourquoi nous étudierons dans le chapitre suivant un système où l'écart moyen est influencé par l'emplacement qu'occupe chaque variante, celle-ci le renforçant proportionnellement à la distance qui le sépare de la moyenne ; c'est la déviation standard.

Quant à la moyenne, quelle est sa valeur en tant qu'élément permettant d'étudier la variabilité ?

Avant tout, pour qu'elle ait sa raison d'être, il est indispensable d'opérer sur un grand nombre de sujets ; il est évident qu'en ne partant que de quelques individus, elle serait considérablement modifiée par une extrême un peu éloignée. De plus, même bien prise, elle n'a pas beaucoup de valeur absolue ; son importance n'est que relative ; il faut encore d'autres facteurs de variabilité car elle ne constitue qu'un élément de base.

DEVIATION STANDARD

Nous avons déjà envisagé deux modes de mesure de la variabilité, l'écart absolu et l'écart moyen et nous avons vu qu'ils étaient loins d'être sans reproches. C'est pourquoi les chercheurs ont voulu trouver des formules plus exactes, plus significatives et surtout moins sujettes à variations. La meilleure d'entre elles est certainement la Déviation Standard.

Avant d'arriver à elle, que Kronacher considère comme la seule mesure de variabilité, d'autres auteurs, entre autres Galton, avaient imaginé le Quartil et son coefficient de variabilité. Pour son calcul il fallait diviser la suite de variations en quarts et c'est la moitié du nombre des fréquences comprises dans les deux quarts du milieu, qui la représentait. Il n'est plus utilisé de nos jours car il dépend trop de la médiane et n'a pas la même valeur que la déviation standard.

Celle-ci, désignée en Biométrie par la lettre σ est la racine carrée de la somme des carrés de l'écart individuel a , multiplié par les fréquences f correspondantes, divisé par le total n des fréquences, et diminué de l'élément correcteur b au carré. On l'obtient donc d'une manière analogue à celle de l'écart moyen sauf que l'on porte l'écart individuel a à la

deuxième puissance a^2 et que l'on fait ensuite l'extraction de la racine carrée du résultat. La formule

$$\text{est donc } \sigma = \pm \sqrt{\frac{\sum f a^2}{n} - b^2}$$

C'est la formule simplifiée de celle, plus complète,

$$\text{donnée par Johannsen : } \pm \sigma = \pm \sqrt{\frac{\sum f a^2}{n-1} - b^2}$$

mais son auteur lui-même a indiqué que pour son utilisation, étant donné qu'elle porte sur des variations d'au moins 50 unités, on peut confondre $n-1$ et n sans amener d'erreur dans les calculs.

En multipliant la déviation standard σ par 100 et en divisant par la moyenne arithmétique on obtient le coefficient de variation

$$\sigma v = \pm \frac{\sigma \cdot 100}{M} \%$$

Comme application pratique, envisageons ces mesures pour la variation de l'écartement des hanches de nos bovins.

h	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49
a	—9	—8	—7	—6	—5	—4	—3	—2	—1	0	+1	+2	+3	+4	+5	+6
f	1		1	1	5	10	7	12	22	21	17	10	8	6	3	2

La moyenne supposée est 43 avec la fréquence 21.

A partir d'elle, et dans les deux sens nous numérotions les écarts a puis nous les portons au carré et enfin nous les multiplions par la fréquence f correspondante.

a	—9	—8	—7	—6	—5	—4	—3	—2	—1	0	1	2	3	4	5	6
a ²	81	64	49	36	25	16	9	4	1	0	1	4	9	16	25	36

$$\Sigma a^2 f = 81 + 49 + 36 + 25,5 + 16,10 + 9,7 + 4,12 + 22 + 17 + 4,10 + 9,8 + 16,6 + 25,3 + 36,2 = 956.$$

$$\frac{\Sigma f a^2}{n} = \frac{956}{126} = 7,58.$$

Pour le calcul de b^2 nous suivrons la même marche que lors de la recherche de la moyenne. Nous établissons le tableau :

Déviatiou a	1	2	3	4	5	6	7	8	9
Fréquence +	17	10	8	6	3	2			
Fréquence —	22	12	7	10	5	1	1		1
Différence	—5	—2	+1	—4	—2	+1	—1		—1
Σ diff. a	—5	—2,2	+3	—4,4	—2,5	+6	—7	—9	= —42.

$$b = \frac{-42}{126} = -0,333 \quad b^2 = 0,11$$

Reprenons le calcul de la déviation Standard.

$$\sigma = \pm \sqrt{\frac{\Sigma f a^2}{n} - b^2} = \pm \sqrt{7,47} = 2,75.$$

Calculons aussi le coefficient de déviation :

$$\frac{\sigma \cdot 100}{M} \%$$

$$\text{La moyenne } M = A + b = 43 - 0,33 = 42,77.$$

$$\sigma V = \frac{2,75 \cdot 100}{42,77} = 6,4 \%$$

Certains auteurs prennent pour F une formule qui, à première vue, semble différente. Zwaenepool pose :

$$\sigma = \pm \sqrt{\frac{\Sigma f a^2}{n}}$$

l'élément correctif b^2 manque ; mais il est déjà sous-

entendu dans a qui représente non plus la déviation à partir de la moyenne supposée A mais à partir de la moyenne arithmétique vraie.

La déviation standard est, avons-nous dit, avec son coefficient, la meilleure mesure de la variabilité. Mais quel est son mode d'utilisation ?

Elle est indispensable pour pouvoir comparer les chiffres trouvés avec les nombres théoriques, calculés : une correspondance entre les deux indique une grande homogénéité, une fixité et par conséquent la pureté d'une population qui prendra alors le nom de race. Ainsi un groupement assez étendu de bovins ayant satisfait à ces conditions, au lieu d'être par exemple le « bétail brun de la région des Alpes » deviendra « la race brune des Alpes ». Alors que la première appellation implique une notion de mélange d'individus qui ont la faculté de vivre dans des conditions géographiques données, la seconde parle de bovins à caractères fixes qui, même sortis de leur habitat primitif, garderont leurs qualités, ce qui n'est pas forcément le cas pour les premiers.

La comparaison entre la théorie et la pratique se fait par deux méthodes différentes : l'une arithmétique, l'autre graphique.

La méthode arithmétique se base sur un rapprochement des chiffres, d'abord quelque peu empirique, puis très mathématique par l'utilisation de certaines tables. Un premier tableau est celui de la « répartition normale ou binomiale » ; il représente le nombre de fréquences sur 10.000 qui se trouvent entre la moyenne — dont l'écart est zéro — et une

valeur d'écart positif ou négatif donnée en valeur standard — c'est-à-dire écart divisé par la déviation standard. Il a été donné par Johannsen dans son livre *Elemente der exakten Erblchkeitslehre*, p. 77 et se trouve reproduit dans la plupart des ouvrages de génétique et de Biométrie.

Le second tableau est plus complexe : c'est une table de probabilités dite « table pour X^2 » — X^2 étant une valeur que nous verrons ultérieurement — ; il comprend pour chaque nombre de classes n une série de valeurs de X^2 , et, en regard, le degré de la probabilité P en centième, l'unité représentant la concordance parfaite. Elle est donnée par Fischer, dans *Statistical Methods for Research Workers*. Il est disposé de la manière suivante :

n	$P = 0,99$	$P = 0,98$	$P = 0,95$	
1	0,000157	0,000628	0,00393	
2	0,0201	0,0406	0,103	

Reprenons la variation des longueurs de croupe :

lc	35	36	37	38	39	40	41	42	43	44	45	46	47	48
n	1	4	10	16	13	17	17	15	16	11	4	3	2	1

Nous répétons les mêmes calculs que précédemment pour la déviation standard et nous obtenons :

$$\sigma = + \sqrt{7,285 - 0,04} = + 2,69.$$

$$M = A - b = 41 - 0,2 = 40,8.$$

A l'aide de la première table de Johannsen, nous établissons le tableau suivant :

1	2	3	4	5	6	7	8
Classes	Limites de classes	Ecart a	a : F	Fréquences théoriques			Fréquences obtenues
				Somme	Pour 10.000	Pour 130	
					par classe		
35	34,5	- 6,3	- 2,34	4903	147	1,91	1
36	35,5	- 5,3	- 1,07	4756	318	4,15	4
37	36,5	- 4,3	- 1,59	4438	551	7,16	10
38	37,5	- 3,3	- 1,22	3887	864	11,23	16
39	38,5	- 2,3	- 0,85	3023	1179	15,33	13
40	39,5	- 1,3	- 0,48	1844	1409	18,25	13
41	40,5	- 0,3	- 0,11	438			
	M=40,8	0	0	0	1464	19,03	17
42	41,5	+ 0,7	0,26	1026	1331	17,30	17
43	42,5	+ 1,7	0,63	2357	1056	13,73	16
44	43,5	+ 2,7	1,00	3413	732	9,52	11
45	44,5	+ 3,7	1,37	4145	445	5,73	4
46	45,5	+ 4,7	1,74	4590	239	3,11	3
47	46,5	+ 5,7	2,12	4329	107	0,49	2
48	47,5	+ 6,7	2,49	4936	043	0,26	1
	48,5	+ 7,7	2,86	4979	9882	132,25	130

Dans la colonne 1 nous portons les variantes, les classes, puis à côté les limites entre lesquelles elles sont comprises. Dans la 3^e colonne nous calculons les écarts entre ces limites et la moyenne arithmétique. Dans la suivante se trouve la valeur $a : \sigma$ obtenue en divisant les chiffres de la colonne 3 par $\sigma = 2,69$; cette valeur est indispensable pour le calcul des fréquences théoriques avec la table de Johannsen : celle-ci est graduée de 0,05 en 0,05 : en regard de chaque valeur de $a : \sigma$ se trouve le chiffre total de fréquences théoriques pour $n = 10.000$, depuis la moyenne pour laquelle $a : \sigma = 0$ jusqu'au chiffre cherché ; puis à côté Johannsen a indiqué la correction à faire pour chaque centième 0,01 en plus. Grâce à cette table nous avons obtenu la colonne 5. Dans la 6 sont indiquées les fréquences, toujours théoriques existant entre les limites précitées* ; pour les connaître il suffit de retrancher l'un de l'autre les nombres totaux des fréquences jusqu'à ces limites ; exemple : à partir de 40,8 cm. (moyenne) jusqu'à 37,5 cm. nous avons une fréquence de 3887, jusqu'à 36,5 cm., 4438, la différence 551 donne la fréquence pour 10.000 comprise entre 36,5 et 37,5 cm., donc pour la variante 37 cm.

Enfin dans la colonne 7 nous ramenons les fréquences théoriques données pour $n = 10.000$ à celles correspondant à $n = 130$, en multipliant la précédente par le rapport $130/10.000$ et nous les comparons aux chiffres de la colonne 8 qui sont ceux qui ont été donnés expérimentalement.

En additionnant les chiffres de la colonne 6 et ceux de la colonne 7 nous devrions obtenir respectivement 10.000 et 130. Nous n'avons que 9882, soit donc une erreur de 118 qui, par rapport à 10.000 est assez faible — 1,1 % —. Pour la seconde le résultat est 132,25, nombre supérieur ; cette deuxième erreur est consécutive à la première car en voulant ramener à 130 une fréquence de 10.000, nous n'avons calculé en fait qu'avec 9882.

Nous pouvons déjà comparer ainsi les chiffres des colonnes 7 et 8 et remarquer une certaine concordance. Mais pour pouvoir juger des résultats obtenus d'une façon plus mathématique, nous allons utiliser la table des probabilités de Fischer, mais auparavant faire les calculs portés sur le tableau suivant.

Entre 1 et 2 nous reportons les colonnes 7 et 8 précédentes. En 3 nous faisons la différence que nous portons au carré en 4. Enfin dans la colonne 5 cette différence entre la théorie et la réalité portée au carré est divisée pour chaque classe par la fréquence qui lui correspond. Le total des chiffres de 5 nous donne la valeur X^2 qui est en quelque sorte un total d'erreurs relatives.

En consultant la table X^2 de Fischer, dans la rangée horizontale 14 — le nombre de classes de variantes est 14 — nous cherchons le nombre le plus approchant de $X^2 = 11,37$; c'est 10,82 correspondant à la rangée verticale de $P = 0,70$.

1	2	3	4	5
Fréq. théoriq. pour 130	Fréq. trouvées pour 130	Différence	Diff. au carré	Colonne 4 divisé par 1
1,91	1	0,91	0,82	0,42
4,15	4	0,15	0,01	0,00
7,16	10	— 2,84	8,06	1,12
11,23	16	— 4,77	21,05	1,77
15,33	13	2,33	5,29	0,34
18,25	17	1,25	1,45	0,07
19,03	17	2,03	4,12	0,21
17,80	15	2,30	5,29	0,3
13,73	16	2,27	1,85	0,13
9,52	11	1,48	2,19	0,23
5,78	4	1,78	3,16	0,54
3,11	3	0,11	0,01	0,00
0,49	2	1,51	2,28	4,64
0,26	1	0,74	0,54	2
				$\chi^2 = 11,37$

On peut dire que la concordance $P = 0,70$ est très bonne. On admet en effet qu'elle existe pour les valeurs de P entre 0,99 et 0,1. La valeur de $P = 0,05$ est considérée comme la limite supérieure de la non-concordance.

C'est maintenant qu'il faudra répéter les remarques que nous avons émises sur nos mensurations lors de l'étude des écarts absolu et moyen : on est bien en présence d'une race fixe et très homogène.

Toutefois nous ferons remarquer qu'il ne faut pas être par trop catégorique et que malgré la rigueur et l'exactitude des calculs il ne convient pas de leur attribuer une valeur qu'ils n'ont pas, car on ne pourra jamais prétendre traduire rigoureusement par des chiffres des données biologiques.

La méthode graphique est le procédé de choix pour réaliser la traduction synthétique d'un phénomène dont elle indique, mieux que des calculs, l'allure générale. A ce point de vue elle constitue un moyen de recherches et même de contrôle. Elle a pour but de comparer la courbe trouvée pratiquement avec une courbe idéale construite d'après les données binomiales.

Construction de la courbe idéale :

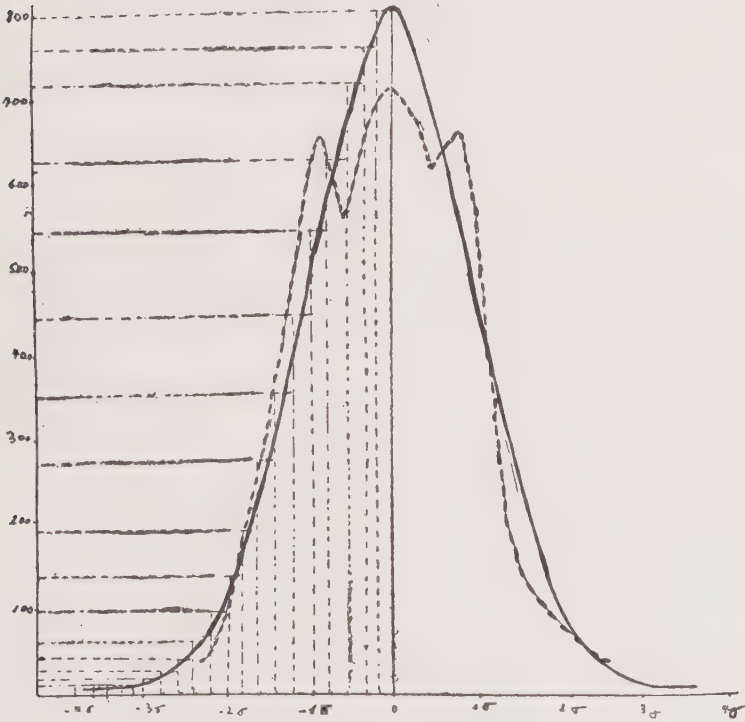
Sur l'axe vertical nous portons les fréquences de chaque classe à l'échelle suivante : 1 mm. pour une fréquence de 5.

Sur l'axe horizontal nous portons les classes, exprimées par leur écart en fonction de $\sigma = 2$ cm. Nos classes auront un écart de $0,2\sigma$ et pour chacune d'elles nous calculerons f à l'aide de la table de Johannsen. Nous obtenons

Classe 1.... 793	Classe 8.... 260	Classe 15.... 13
— 2.... 761	— 9.... 189	— 16.... 6
— 3.... 704	— 10.... 132	— 17.... 4
— 4.... 623	— 11.... 88	— 18.... 1
— 5.... 532	— 12.... 57	— 19.... 1
— 6.... 436	— 13.... 35	— 20.... 1
— 7.... 343	— 14.... 21	

La représentation graphique donne la courbe en trait plein. Chaque classe est représentée par un rectangle dont la hauteur est la fréquence et la base est l'écartement de ces limites $0,2$. C'est en réunissant les milieux des sommets qu'on obtient la courbe

cherchée. Sur la moitié gauche du schéma nous avons établi la courbe théorique et nous en avons établi la symétrique à droite. Puis nous avons établi la courbe des chiffres relevés — en pointillé — Pour ne pas surcharger la figure nous n'avons donné que le résultat.



Pour pouvoir placer les chiffres relevés expérimentalement sur les mêmes unités il est nécessaire de leur faire subir quelques modifications. Les fréquences seront d'abord reportées comme les théoriques à $n = 10.000$; ils seront donc multipliés par $\frac{10.000}{130}$

D'autre part, l'écart entre les classes du grand tableau précédent, colonne 4, est $0,37 \sigma$. Il faudra donc modifier le chiffre une seconde fois pour obtenir l'ordonnée du rectangle ; les fréquences seront donc

multipliées par $\frac{0,2}{0,37}$.

En somme les chiffres trouvés (colonne 8) seront multipliés par $\frac{10.000}{130}$ puis par $\frac{0,2}{0,37}$ soit par

$$\frac{10.000 \times 0,2}{130 \times 0,37} = \frac{1.000}{24} = 41,6$$

et nous obtenons la nouvelle variation, dérivée de la première. Les écarts de classe sont donnés par la colonne 4 en fonction de σ .

Ecarts..	— $2,34 \sigma$	— $1,97 \sigma$	— $1,59 \sigma$	— $1,22 \sigma$	— $0,85 \sigma$	— $0,48 \sigma$
Fréq.	42	166	416	666	541	

Ecarts	— $0,48 \sigma$	— $0,11 \sigma$	+ $0,26 \sigma$	+ $0,63 \sigma$	+ 1σ	+ $1,37 \sigma$
Fréq...		707	707	624	666	458

Ecarts	+ $1,37 \sigma$	+ $1,74 \sigma$	+ $2,12 \sigma$	+ $2,49 \sigma$	+ $2,86 \sigma$
Fréq...	166	125	83	42	

Nous portons sur le graphique précédent les rectangles ayant pour base la différence de deux chiffres d'écart à la suite et pour hauteur la fréquence correspondante et nous unissons les milieux des sommets ; la courbe obtenue est la courbe empirique ou réelle.

Elle est irrégulière et présente deux petits sommets accessoires ; mais dans sa forme générale et surtout par la surface qu'elle occupe, elle correspond

bien à la courbe théorique. Or cette surface représente le chiffre total des fréquences : en multipliant la base par la hauteur de chaque rectangle nous refaisons le même calcul qu'à la page précédente mais en sens inverse ; en additionnant les rectangles on obtient le total 10.000.

ERREUR MOYENNE

Toutes les dimensions que nous avons prises ont été, dans une certaine mesure, entachées d'erreur : cela tient à deux causes : la première est due à des fautes d'ordre subjectif, la seconde au fait que la technique de mesure est assez grossière. C'est d'ailleurs ce qui nous a conduit à répartir nos résultats en classes bien qu'il y ait pour chacune une mesure propre et exacte. D'autre part, les individus observés, bien qu'assez nombreux, ne constituent qu'une partie minime de la totalité de la race en question, on n'a jamais pu mesurer ou peser une espèce domestique de son premier à son dernier représentant. Et même si, dans un cadre de descendance très étroit, on avait étudié tous les cas — ce qui a souvent été fait — ils ne représentent tout de même qu'une partie de la postérité possible. Enfin, même dans des circonstances de vie se rapprochant le plus possible, une suite de variations ne comprendra toujours qu'une coupure, un « échantillon » d'une autre suite, bien plus riche en réalités et en possibilités.

Par ces quelques considérations on pourra constater une certaine carence de toutes les valeurs reposant sur des mensurations, telle la moyenne qui devrait être l'expression d'une propriété idéale et typique des individus envisagés.

Pour chacune des moyennes d'une suite de variations il faudra se poser la question suivante : à quel degré peut-on considérer la moyenne calculée comme expression de la propriété ou caractéristique envisagée ?

La réponse est donnée par la formule de « l'erreur probable moyenne » ou « erreur moyenne », $m = \pm \sigma : \sqrt{n}$.

L'erreur moyenne donne le moyen de vérifier la certitude d'une moyenne. Comme nous le montre l'expression précédente, elle n'est proportionnelle qu'à la racine carrée du nombre d'observations.

Reprenons comme exemple la variation de h , déjà exposée lors de la déviation standard. Cette dernière est $+ 2,75$; $n = 126$. D'où l'erreur moyenne est : $+ 2,75 : 1,12 = + 2,5$.

La moyenne rectifiée par cette erreur est $M \pm m = 42,7 \pm 2,5$.

Avant d'envisager l'application pratique de la mesure des erreurs, il est nécessaire de faire les observations suivantes :

Chaque suite de variations ne constitue qu'une partie de la suite absolue, totale. Les moyennes obtenues pour les différentes expériences sont donc plus ou moins éloignées de celles que nous appellerons la « moyenne générale » ; elles se groupent autour de celle-ci selon des lois du hasard ; leur représentation graphique donne une courbe. Or les mathématiciens ont calculé que la déviation standard de cette courbe est comprimée par la valeur $\sigma : \sqrt{n}$, σ et n se rapportant aux séries de variations prises. La courbe

construite à partir de cette déviation standard se rapproche d'autant plus de la courbe normale des erreurs que n est plus grand.

Pour une courbe normale on peut calculer le nombre d'individus qui se trouvent à droite ou à gauche de la moyenne et à des écarts donnés 1σ , 2σ ou 3σ . Dans l'espace compris entre -3σ et $+3\sigma$ on trouve 95,74 % du total ; pour l'écart $+2\sigma$ on ne trouve que 95,46 %. On peut donc dire d'une variante quelconque qu'elle a 99,74 % de chances de se trouver à 3σ de la moyenne générale.

Inversement, la moyenne générale se trouve dans ces limites à partir de la moyenne calculée dans la variation en question. La répartition des moyennes partielles possibles se fait sur une distance $m = \sigma : \sqrt{n}$; l'espace pour la moyenne générale est donc de $\pm 3m$ ou $\pm 2m$ suivant la probabilité désirée. Johannsen calcule avec $\pm 3m$ tandis que Fischer se tient à $\pm 2m$ et d'autres, qui veulent plus de certitude encore, calculent avec $+4m$, voire même $\pm 5m$. Pour les recherches biométriques de zootechnie $\pm 3m$ est bien suffisant.

La moyenne générale pour l'écartement des hanches de ces limousins âgés de 18 mois est donc comprise entre les limites $M = 42,7 \pm 3 \times 2,5$, soit entre 35,2 cm. et 50,2 cm.

L'erreur probable moyenne nous servira aussi à vérifier la valeur d'utilisation de cette moyenne en tant que base de mesure. Pour cela elle devra être au minimum le triple de la valeur de son erreur. Si ce n'est pas le cas, la limite inférieure $-3m$ se-

rait négative, ce qui est un non-sens et empêche l'emploi de ce résultat.

Il arrive aussi qu'une autre question se pose : après avoir établi deux séries de variations sur la même espèce, mais soit dans deux contrées différentes soit dans des circonstances différentes, on pourra se demander si ces deux parties appartiennent à la même et unique variation générale de l'espèce ; ou bien s'agit-il de deux populations différentes ?

Pour s'en rendre compte le calcul est très simple grâce à cette erreur moyenne : on recherche d'abord la différence entre les moyennes $d = M_1 - M_2$ puis l'erreur de la différence des deux moyennes qui est donnée par la formule :

$$m \text{ (diff.)} = \sqrt{m_1^2 + m_2^2}.$$

m_1 et m_2 représentent les deux erreurs moyennes. Si la différence des moyennes est supérieure à trois fois l'erreur de la différence, tous les individus étudiés appartiennent à une même suite de variations ; mais si $d < 3 m \text{ (diff.)}$ on est en présence de groupements différents.

Cette erreur moyenne constitue donc non seulement une mesure de l'exactitude de nos observations mais encore une mesure de variabilité. Car plus celle-ci est grande pour une dimension, plus l'écart en dedans duquel se trouve la moyenne générale, est grand ; et, moins une race est fixe dans certaines qualités, moins cette qualité pourra être considérée comme typique. L'hérédité et l'intensité de la transmission d'une donnée considérée comme caractéristique se trouve encore en relation très étroite avec

la variabilité ; c'est-à-dire que celle-ci est conditionnée pour de mêmes inconstances extérieures par la prédisposition plus ou moins régulière à ce caractère. Pour une erreur moyenne plus petite on peut aussi admettre, toujours pour les mêmes circonstances extérieures, une tendance à une plus grande puissance de transmission, à condition, naturellement, d'opérer sur un nombre d'observations identiques ou au moins approchant. L'erreur moyenne nous permet enfin de reconnaître le progrès d'unification subie par une race, soit par la sélection, qui est l'œuvre du zootechnicien, soit par l'ensemble des conditions vitales, œuvre de la nature ; pour cela il suffit de prendre des mensurations périodiques et de comparer.

La formule que nous avons donnée pour l'erreur moyenne est celle utilisée par Fischer, Johannsen, Kronacker. Certains auteurs, entre autres Zwaenepool, préfèrent utiliser la valeur $m = 0,67 \sigma : \sqrt{n}$. Ils introduisent un facteur supplémentaire constant qui est 0,67 — représentant la valeur approchée de $0,666... = 2/3$. Or $2/3 \sigma$ représente assez approximativement la valeur du quartil calculé.

La première formule nous a semblé préférable parce que plus sévère : l'erreur moyenne obtenue à partir d'elle est plus élevée que l'autre et donne ainsi une plus grande marge de sécurité.

Dans une suite de variations on peut envisager encore d'autres erreurs probables. L'erreur probable d'une série de variantes ou erreur moyenne étant $\sigma : \sqrt{n}$, lorsque n sera réduit à l'unité, on aura l'er-

reur probable d'une variante isolée qui sera σ' , la déviation standard. Pour un groupe quelconque de variantes, l'erreur probable sera aussitôt donnée par la même formule.

Enfin il existe encore une erreur probable dans la déviation standard qui est $m(\sigma') = \frac{\sigma}{\sqrt{2n}}$; elle est inversement proportionnelle à la racine carrée du double des fréquences.

Mais toutes ces dernières erreurs n'ont pas la même valeur que l'erreur moyenne, qui, elle, peut servir, à un certain degré, de mesure de variabilité.

LA CORRELATION

De toutes les méthodes que nous avons envisagées jusqu'à présent pour l'étude des mensurations, il ne s'agissait que de l'examen d'une donnée unique et la variation ne concernait qu'une propriété.

Pour l'explication de nombreux problèmes biologiques on est parfois conduit à rechercher si deux propriétés peuvent avoir dans leur variabilité un rapport entre elles et surtout dans quelle mesure. Ce rapport est désigné sous le nom de corrélation.

Il est tout naturel que dans l'organisme animal de nombreuses parties se trouvent dans une certaine dépendance les unes des autres : les lois du balancement des organes ont été établies depuis longtemps par Cuvier, Darwin et Baron et forment l'essence de nos doctrines constitutionnelles actuelles. Mais depuis ces noms illustres, des progrès ont été faits et on a cherché ces derniers temps des corrélations entre des données des plus éloignées en apparence. C'est ainsi que Duerst a établi un rapport d'ailleurs constaté entre l'angle costal de la vache et la production laitière.

Faisons remarquer en passant que le terme de corrélation peut conduire à une méprise car l'esprit tend à lui donner une signification de causalité : c'est

ce que fait remarquer M. Julin dans les *Principes de statistique théorique appliquée* : « En logique, le caractère corrélatif de deux phénomènes implique l'existence entre eux d'un lien causal » et certains auteurs qui se sont attachés au côté mathématique de la question plutôt qu'à son aspect philosophique, n'ont pas hésité à écrire que l'existence d'une corrélation d'un degré assez élevé entre deux phénomènes, supposait l'existence de causalité.

C'est pour cette raison que nous aussi nous nous rangeons à l'opinion de L. March qui juge le terme de co-variation plus exact.

Il ne faudrait pas passer sous silence la belle définition très générale donnée par Viola : « L'étude des corrélations est représentée par une étude exacte de l'architecture externe de l'organisme, afin de l'envisager en soi et pour soi et de l'interpréter en tout ce qu'elle peut nous servir pour nous éclairer sur l'architecture interne. »

Cette explication, qui nous ouvre des horizons très vastes, nous entraîne à chercher une classification de toutes les corrélations possibles ; leur étude est basée sur la nature des caractères corrélés et comprend les six classes suivantes :

1° Etude des variations corrélatives de l'architecture externe, ou éso-extérieures ; elle intéresse directement l'Extérieur et l'Anthropométrie chez l'homme ; c'est l'expression mathématique d'une plus ou moins grande harmonie esthétique de la race : elle sera donc d'une grande valeur pour les animaux de travail ou même de boucherie.

2° Etude des variations corrélatives entre architectures externe et interne, ou eso-internes : cette partie présente un intérêt relativement faible; elle envisage par exemple le rapport entre les dimensions de l'appareil digestif et de la structure externe.

3° Etude des variations corrélatives entre l'architecture externe et les fonctions organiques ou eso-fonctionnelles, très importante, car elle se confond avec le but de la zootechnie qui est d'obtenir le rendement optimum des productions animales. Nous comparerons ultérieurement cette méthode d'étude théorique avec celle, pratique, donnée par le contrôle laitier par exemple.

4° Etude des variations de l'architecture interne, ou endo-internes; très utiles pour l'anatomiste, l'embryologiste, l'histologiste et, bien que rarement, pour le zootechnicien.

5° Etude des variations corrélatives entre l'architecture interne et les fonctions ou endo-fonctionnelles. C'est cette étude qui a permis à Bichat de découvrir d'une façon merveilleuse les fonctions antitoxique et glycogénétique du foie à partir de considérations uniquement morphologiques : il a considéré le poids du foie comparativement à la masse totale des autres glandes de l'organisme et en a conclu que l'excrétion biliaire, si minime si on la compare aux autres sécrétions, ne pouvait pas être la seule fonction du foie.

6° Etude des variations corrélatives entre deux fonctions ou corrélations physiques fonctionnelles. C'est là un domaine uniquement physiologique et

auquel on peut reprocher d'utiliser presque exclusivement les méthodes qualitatives.

Tous ces différents chapitres ouvrent un champ d'étude embrassant absolument toutes les sciences s'appliquant aux animaux. Mais pour nous, qui sommes avant tout des zootechniciens, seules les trois premières parties nous intéressent. Nous avons déjà signalé en passant que l'étude des corrélations éso-internes est d'un intérêt relativement faible. Celle des corrélations éso-fonctionnelles sont très importantes pour l'amélioration du cheptel en vue des productions animales; mais c'est surtout à la première que nous allons nous adresser.

Les corrélations éso-extérieures ont été les premières étudiées parce que les plus faciles à établir. En effet, pour pouvoir chercher une corrélation éso-fonctionnelle, par exemple, il est nécessaire de pouvoir suivre de nombreux sujets très longtemps, donc de disposer d'un champ d'expériences parfait. C'est pourquoi, c'est en partant de nos premières mesures, toutes externes, que nous allons envisager cette petite étude critique.

Pour juger de la valeur d'une corrélation supposée nous avons à notre disposition deux moyens : d'une part le calcul du coefficient de corrélation, d'autre part l'établissement du tableau de corrélation.

Le coefficient de corrélation est donné par la formule de Bravais :

$$r = \frac{\Sigma p.a_x . a_y - n b_x b_y}{n. b_x . b_y}$$

dérivée de la vraie formule de corrélation, moins pratique à utiliser :

$$r = \frac{\sum p \cdot \alpha_x \cdot \alpha_y}{n \cdot \sigma_x \sigma_y}$$

Comme dans les formules précédentes :

α représente l'écartement de la moyenne vraie M.
 a représente l'écartement de la moyenne supposée A.

X représente la propriété de base, celle à laquelle on compose l'autre.

Y représente la propriété comparée.

σ représente la déviation standard.

on comprendra aisément pourquoi on a adopté la première formule, bien que d'apparence plus complexe, mais il est bien plus facile de calculer à partir d'une moyenne supposée A que de la moyenne vraie M, quitte à apporter dans la suite le correctif :

$$= \frac{n b_x b_r}{n \sigma_x \sigma_r}$$

Ce coefficient de corrélation est indépendant de l'unité de mesure des deux caractères, car il n'est en quelque sorte qu'un numéro, les écarts individuels de l'un et de l'autre caractère exerçant de l'influence sur le numérateur comme sur le dénominateur, ce qui permet toujours de comparer la valeur du coefficient obtenu pour deux séries avec celle obtenue pour deux séries différentes.

Ce coefficient varie de +1 à —1; +1 exprime la corrélation positive parfaite et —1 la corrélation négative absolue; toutes les valeurs intermédiaires sont possibles et plus il s'approche de l'unité, plus la cor-

rélation est parfaite; la corrélation est positive lorsque les deux caractères varient dans le même sens, sinon elle est négative.

Dans le calcul de ce coefficient intervient aussi une erreur par suite de la répartition irrégulièrement binomiale des variantes. Aussi est-il bon de considérer l'erreur probable du coefficient de corrélation r ; cette erreur moyenne est donnée par la formule de Pearson :

$$m(r) = \pm \frac{1 - r^2}{\sqrt{n}}$$

Avec Mortara on admet qu'il faut exiger que le coefficient r soit quatre ou cinq fois plus grand que son erreur probable.

La seconde méthode, moins mathématique et plus schématique, est l'établissement des tables de corrélations.

On commence, comme auparavant, par répartir le matériel par classes, on établit la suite de variations pour chaque propriété. L'une des deux propriétés, n'importe laquelle, est nommée caractère X ou propriété de base, c'est à elle qu'on compare l'autre. Cette dernière, Y , doit être mesurée indépendamment du caractère X ; c'est le caractère relatif ou comparé.

On porte les classes du caractère de base X dans la rangée verticale de gauche, en ordonnées, et en abscisses, en haut celles de Y . Puis on recherche de quelle manière les fréquences de X se répartissent sur les classes des autres fréquences de Y , c'est-à-dire

à quelle classe de X appartiennent les variantes de Y. Nous avons une corrélation égale à ± 1 lorsque la ligne obtenue en réunissant par la pensée les chiffres inscrits, sera droite et unique, s'étendant en diagonale d'un coin à l'autre du rectangle; cette corrélation, en ligne droite, n'existe pas en Biométrie appliquée; toutes les représentations que nous pourrions obtenir ne pourront que l'approcher. Parfois même les variantes seront irrégulièrement réparties sur tout le tableau, nous dirons alors que la corrélation n'existe pas.

A titre d'exemple nous allons envisager la valeur de la corrélation entre la longueur de tête L et la largeur l chez les mâles. Nous obtenons le tableau suivant :

	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	
16					I												I
17	I		I	4	I	I		I									9
18			I	2	4	5	8	4	I								25
19					5	4	5	9	5	3	I					I	33
20					I	I	2	7	4	4		I			I		21
21										I	I	I					3
22											I						I
	I		2	6	12	11	15	21	10	8	3	2			I	I	93

Sur la ligne horizontale supérieure nous avons la suite des classes $Y = L$; la colonne gauche extrême est la suite $X = l$. On verra qu'en additionnant les chiffres de chaque ligne horizontale on obtient la variation de X qui est représentée par la colonne de

droite; il en est de même de la variation du Y qui est représentée par la dernière ligne.

Ce tableau, dont les chiffres sont répartis d'une manière à peu près rectiligne du coin droit supérieur au coin gauche inférieur rappelle assez bien une corrélation positive, droite.

Pour pouvoir calculer le degré de corrélation d'après la formule de Bravais, on modifie légèrement la présentation des ordonnées et abscisses. Comme pour le calcul de la moyenne et de la déviation standard on utilise les valeurs moyennes supposées A et A et les écartements a et a positifs à droite ou en bas, négatifs à gauche ou en haut.

	-7	-6	-5	-4	-3	-2	-1	$\frac{Ay+11}{0}$	+1	+2	+3	+4	+5	+6	+7	+8	
-3	+				I ₉											-	I
-2	I ₁₀		I ₁₀	4 ₁	I ₈	I ₄		I ₀									9
-1			I ₃	2 ₁	4 ₁	5 ₁	8 ₁	4 ₀	I ₁								25
0					5 ₀	4 ₀	5 ₀	9 ₀	5 ₀	3 ₀	I ₀					I ₀	33
+1					I ₃	I ₁	2 ₁	7 ₀	4 ₁	4 ₁		I ₄			I ₁		21
+2										I ₄	I ₆	I ₃					3
+3	-										I ₉					+	I
	I		2	6	I2	11	I5	21	10	8	3	2			I	I	

Grâce à cette division croisée le tableau se trouve partagé en quatre quartiers : le quartier inférieur droit donne les valeurs positives de X et de Y; à gauche, en haut, ce sont les valeurs négatives de X et de Y; à droite en haut les valeurs positives de Y et négatives de X, à gauche en bas les valeurs négatives de Y et positives de X.

Les chiffres portés dans le quadrillage indiquent le nombre de fois qu'à une valeur donnée d'une propriété X correspond une autre valeur fixe de Y. Les chiffres portés en indice représentent le produit $a_x \cdot a_y$, c'est-à-dire le produit de l'écart entre les moyennes $A_x = 19$ cm. et $A_y = 44$ cm. et les classes (ordonnées et abscisses) du chiffre.

Les quartiers et par conséquent ses coefficients ont différents signes :

à gauche en haut — . — = +

à droite en haut — . + = —

à gauche en bas + . — = —

à droite en bas + . + = +

Dans chaque quartier les chiffres de fréquence seront multipliés par les indices correspondants puis affectés du signe propre à chaque quart — signe qui est porté à chaque coin; on additionne les positifs et on soustrait les négatifs.

$$\begin{array}{l} \text{à g. en haut} \\ \Sigma p a_x a_y = + 118 \end{array}$$

$$\begin{array}{l} \text{à dr. en bas} \\ \Sigma p a_x a_y = + 50 \end{array}$$

$$\begin{array}{l} \text{à g. en bas} \\ \Sigma p a_x a_y = - 7 \end{array}$$

$$\begin{array}{l} \text{à dr. en haut} \\ \Sigma p a_x a_y = - 1 \end{array}$$

La somme totale

$$\Sigma p a_x a_y = 118 + 50 - (7 + 1) = 160$$

Les autres chiffres nécessaires pour le calcul du numérateur, à savoir $n b_x b_y$ sont calculés suivant la méthode déjà utilisée:

Calcul de b_x :

a	0	1	2	3	n=33
+	(33)	21	3	1	
—		25	9	1	$A_x = 13$
Diff.		— 4	— 6		

Déviation négatives : $-4.1 + -6.2 = -6$.

$$\text{D'où } b_x = \frac{-16}{93} = -0,172$$

	a	0	1	2	3	4	5	6	7	8
Calcul de b_y	+	(27)	10	8	3	2			1	1
n = 93 $A_y = 44$	—		15	11	12	6	2		1	1
Diff		—5	—3	—9	—4	—2				+1

Déviation positives : $1.8 = +8$,

Déviation négatives : $(5.1 + 3.2 + 9.3 + 4.4 + 2.5 = -64$

$$\text{D'où } b_y = \frac{-64 + 8}{93} = -0,602$$

D'où le numérateur

$$\Sigma p a_x a_y - n b_x b_y = 160 - 93.0,172.0,602 = 150,33$$

Les chiffres du dénominateur seront obtenus de la manière déjà développée, calcul de la dérivation standard σ —

Calcul de σ_x	a	0	1	2	3
	+	(33)	21	3	1
	—		25	9	1
Somme			46	12	2

$$\Sigma p a_x^2 = 46 + 12.4 + 2.3 = 112$$

$$\frac{\Sigma p a_x^2}{n} = \frac{112}{93} = 1,2$$

$$b_x = -0,172 \quad b^2_x = 0,3$$

$$\sigma_x = \pm \sqrt{\frac{\sum p a^2}{n} - b^2} = \sqrt{0,9} = \pm 0,948$$

Calcul de σ_y

a	0	1	2	3	4	5	6	7	8
+	(21)	10	8	3	2			1	1
—		15	11	12	6	2		1	
Somme		25	19	15	8	2		2	1

$$\sum p a^2_y = 25 + 13,4 + 15,9 + 8,16 + 2,25 + 2,49 + 1,64 = 576.$$

$$\frac{\sum p a^2_y}{n} = 6,19$$

$$b_y = 0,602 \quad b^2_y = 0,36$$

$$\text{D'où } \sigma_y = \sqrt{6,13 - 0,36} = 2,41.$$

Nous avons maintenant tous les termes nécessaires au calcul de la corrélation d'après la formule de Bravais.

$$r = \frac{\sum p a_x a_y - n b_x b_y}{n \sigma_x \sigma_y}$$

$$r = \frac{+150,33}{93 \cdot 0,948 \cdot 2,41} = +0,723$$

L'erreur moyenne calculée d'après la formule de Pearson est :

$$mr : \frac{1-r^2}{\sqrt{n}} = \frac{1-0,52}{9,64} = 0,049$$

D'après Mortara qui exige que le coefficient de corrélation soit quatre à cinq fois plus grand que l'erreur probable, la valeur de $r=0,723$ peut être considérée

comme bonne, puisque l'erreur 0,049 est 14,5 fois plus petite.

Le degré de la corrélation que nous n'avons fait que constater à l'aide du tableau précédent est de 0,723; ce chiffre signifie qu'il existe des rapports réciproques dans la race envisagée entre la longueur et la largeur de la tête, rapport qu'on pourrait être tenté de traduire par un chiffre de pourcentage. Mais ce rapport n'est pas un rapport absolu, car la longueur de tête n'influe pas directement sur la largeur, ni inversement. Il s'agit ici plutôt d'un rapport relatif à une certaine hérédité. Cette hérédité de croissance de tout le corps et la formation isolée de ses parties est telle que, soumis à de mêmes conditions extérieures, les différentes données, ici largeur et longueur de tête se correspondent dans un mode déterminé, à l'intérieur de limites d'oscillations données. Dans notre cas, les extrêmes sont les limites de variabilité, 37 cm. et 52 cm. pour L et 16 cm. et 22 cm. pour l.

Envisageons d'autres corrélations : en existe-t-il entre la longueur de croupe et l'écartement des hanches : le tableau, bien que assez diffus, semble indiquer une très vague corrélation droite, mais il n'en est rien ; en effet, le calcul nous donne $r=0,230$, donc un coefficient très faible; de plus, l'erreur moyenne se monte à 0,065, elle est trop forte pour que nous puissions en tenir compte.

$$\Sigma p. a_x, a_y = 221.$$

$$b_x = -0,333 ; b_y = -0,206.$$

$$n b_x, b_y = + 8,316.$$

$$\sigma_x = + 2,75 \sigma_y = + 2,79.$$

$$\text{D'où : } r = \frac{221 - 8,316}{966,42} = 0,230$$

$$mr = \frac{1 - r^2}{\sqrt{n}} = 0,063.$$

Cette erreur moyenne m a le grand avantage d'annuler les résultats lors d'une valeur trop faible de n , ce qui est le cas ici.

Enfin, en étudiant les rapports entre la hauteur au garrot et l'écartement des ischiurs, mais uniquement chez les mâles, nous obtenons un coefficient $r = 0,585$ pour une erreur $mr = 0,068$, résultat très acceptable. Nous en concluons qu'il y a une certaine corrélation entre la hauteur au garrot et l'écartement des ischiurs, bien qu'il n'y ait aucune action directe de l'une sur l'autre.

Puisque ce coefficient de corrélation ne peut nous renseigner exactement au sujet de l'existence d'un lien causal entre deux phénomènes que l'on compare, quelle est donc au juste sa fonction et sa signification?

En premier lieu l'indice de corrélation précise le degré de parallélisme ou d'anti-parallélisme des courbes. Il ne suffit pas de constater que les séries sont formées de nombres dont la succession est à peu près la même de part et d'autre, ou en un sens opposé, il faut encore préciser ces ressemblances ou ces divergences. Le coefficient de corrélation en donne l'expression la plus exacte; il permet l'orientation de l'esprit du chercheur qui hésite entre deux hypothèses possibles : y a-t-il un rapport entre la présence simultanée de deux faits, et l'apparition d'un phéno-

mène A provoquera-t-elle inmanquablement l'apparition d'un phénomène B? Le coefficient permet de donner une orientation d'autant plus sûre que sa valeur est plus élevée; de même inversement une valeur trop faible pourra nous dissuader d'associer trop intimement deux séries de faits que nous pourrions trouver voisins.

Enfin, il a un rôle à jouer dans la vérification des raisonnements.

En résumé, nous citerons l'opinion de L. March :

« Dans l'enchaînement des opérations logiques de l'esprit, l'intervention des procédés mathématiques dans les comparaisons statistiques a une utilité évidente aux extrémités de la chaîne; au point de départ pour suggérer les conclusions possibles, ou à la fin pour contrôler la partie analytique du processus ; elle ne peut, en aucune façon, constituer le nœud même du raisonnement. »

Cette dernière affirmation a son importance; les formules mathématiques ne dispensent pas de raisonner; lorsque la valeur de r est déterminée, on n'a, répétons-le, qu'un guide pour l'esprit au début et à la fin de la recherche scientifique. On peut même aller plus loin et dire, avec M. Udny Yule que « la grandeur
« du coefficient de corrélation montre l'amplitude de
« l'écart à attendre dans un phénomène donné, par
« rapport à la moyenne pour un certain écart constaté dans un phénomène connexe, à condition qu'on
« puisse juger de l'avenir d'après les données du
« passé. »

Mais il ne faudrait pas attribuer une valeur trop

absolue au résultat obtenu et vouloir chercher à tout prix une corrélation lorsque la théorie permet de l'admettre. Le hasard joue quelquefois, surtout lorsqu'on s'adresse à un nombre insuffisant de sujets; ainsi, si nous trouvions un rapport entre la longueur de la tête et la production laitière, que faudrait-il conclure? Il faudrait sans aucun doute annuler le résultat, car la logique et la science ne peuvent l'admettre et l'on ne peut trouver ni une action directe ou indirecte de l'une sur l'autre, ni un point commun. Par contre, il en est tout autrement si nous envisageons au lieu de la longueur de tête, une donnée qui serait l'expression générale d'un développement et d'une conformation. C'est d'ailleurs en se basant sur ce principe que les propriétaires d'étables distinguent, à l'achat, les bonnes des mauvaises laitières, mais d'une façon empirique pouvant amener des erreurs.

Cette relation peut s'expliquer par l'action de glandes à sécrétion interne qui sont en équilibre organique et qui, d'une part, déterminent le développement général et d'autre part modifient la sécrétion lactée. C'est pourquoi on trouve une conformation si spéciale des grandes laitières. La recherche de la donnée qui traduirait cette allure générale donne lieu à de nombreuses discussions. Avec M. Duerst, nous admettons que l'angle costal la traduit de manière suffisante.

Il en est de même de la corrélation entre la longueur corporelle, la profondeur de poitrine, la largeur des reins du mouton et la production lainière annuelle : une seule de ces données ne permettra au-

cune conclusion alors qu'il en serait autrement pour un terme résumant tous ceux précités.

Nous ne pouvons passer sous silence l'aide qu'a apportée la connaissance du coefficient de corrélation au début de la génétique. Le point de départ de la théorie darvinienne de l'évolution est surtout dans l'existence des différences entre les individus membres d'une même race ou d'une même espèce, différences souvent passées sous silence par le morphologiste.

Plus tard, c'est grâce à la mesure de la corrélation entre les caractères quantitatifs des parents et ceux des fils que Gaston est parvenu à la loi de la régression filiale, puis plus tard, à celle de l'hérédité ancestrale.

Actuellement, on juge que la basse corrélation entre un caractère des pères et celui des fils est en faveur de l'homogénéité de la race et qu'en même temps une haute corrélation dans une population témoigne de l'hérédité d'un caractère. Ce que, pour conclure, nous résumerons en rappelant ces mots de Boldrini:

« La recherche biométrique peut dire beaucoup sur le problème de l'hérédité et de l'évolution. »

CONCLUSIONS

La base de la Biométrie est constituée par des suites de variations qui, lorsqu'elles répondent à certaines conditions, servent à juger de la valeur du génotype d'une race.

I. — L'Ecart Absolu ne présente pas grand intérêt, pas plus que les limites de variabilité qui n'en sont jamais les vrais termes.

II. — La Moyenne Arithmétique Vraie constitue la base de la mesure de la variabilité, mais à elle seule n'a que peu de valeur.

L'Ecart Moyen contribue un peu à l'augmenter, car grâce à lui, on peut fixer les variantes probables supérieures et inférieures, points de repère appréciables.

III. — La vraie mesure est la Déviation Standard. Elle permet par voie arithmétique — à l'aide de tables — ou par voie graphique, la comparaison la plus juste avec la variation idéale.

IV. -- Toutes ces valeurs sont entachées d'erreur. L'Erreur Moyenne Probable permet de vérifier l'exactitude des chiffres précédents et même de rechercher une parenté entre deux populations.

V. — La Corrélation est exprimée par un chiffre, celui de son Coefficient. Il permet, mieux qu'un tableau, de comparer deux variations entre elles. Lorsqu'il est élevé, et que son erreur est assez faible, il ne faut pas l'accepter aveuglément, car il ne constitue qu'un guide pour l'esprit et non une preuve. L'esprit critique garde tous ses droits.

Enfin, tous ces éléments de Biométrie ne sont applicables qu'à une population fixe et non en voie de sélection; ils nous sont d'un grand secours pour l'étude de l'hérédité et de l'évolution.

<i>Vu : Le Directeur</i>	<i>Le Professeur</i>
<i>de l'Ecole Vétérinaire de Lyon,</i>	<i>de l'Ecole Vétérinaire,</i>
L. JUNG.	Dr DOUVILLE.

<i>Le Doyen,</i>	<i>Le Président de la Thèse,</i>
JEAN LEPINE.	ARLOING.

Vu et permis d'imprimer :

LE RECTEUR, PRÉSIDENT DU CONSEIL DE L'UNIVERSITÉ,
A. LIRONDELLE.

BIBLIOGRAPHIE

- BLARINGHEM. — Problèmes de l'hérédité expérimentale. Paris, 1897.
- BRAVAIS (A.). — Analyses mathématiques sur les probabilités des erreurs de situation d'un point. *Acad. des Sciences : Mémoires présentés par divers savants*, 2^e série, t. IX, 1846.
- DUERST. — Gröndlage der Rinderzücht. Springer, Berlin, 1931.
- FAIVRE. — La variabilité et ses limites. Paris, 1868.
- FRATEUR (D^r J.-Léopold). — Les bases génétiques de la fondation des races animales domestiques, dans *Neue Forschungen im Tierzüchtlehre*, Bern, 1936; *Verbandsdruckrei A. G.*, Bern.
- JULIN. — Principes de statistiques théoriques et appliquées. Ed. Marcel Rivière, Paris, 1921.
- KRONACKER (D^r C.). — Genetik und Tierzüchtungen, Berlin, 1934.
- KRONACKER (D^r C.) & PATOW (D^r Frhr V.). — Biometrik, Berlin, 1930. *Verlagbrückhandlung*, Paul Parey.
- MARCO (D^r) & LAHAYE (D^r J.). — Génétique animale, Paris, 1932. *Gemblouse*, Jules Duculot, éditeur.
- USUELLI. (Prof. D^r Philipppo). — Recherches sur les variations corrélatives, dans *Neue Forschungen im Tierzücht und Abstammungslehre*, Bern, 1936.
- ZWAENEPOOL (Prof. D^r H.). — Eléments de Zootechnie générale et de génétique animale. Bruxelles, 1922.

TABLE DES MATIERES

Avant-Propos.	7
Introduction.	9
La Variabilité.	13
Déviatiion standard	31
Erreur moyenne	45
La corrélation	51
Conclusions.	67
Bibliographie.	69



BOSC FRÈRES, M. & L. RIOU

°° °° °° IMPRIMEURS-ÉDITEURS °° °° °°

°° 42, QUAI GAILLETON, 42 °°

°° °° °° °° LYON °° °° °° °°



